<p style="text-align: center;">**REMARKS**</p>

Reconsideration and allowance of the claims are requested in view of the above amendments and the following remarks. Claims 1, 13, 15-17, 22, 25, 27 and 29 have been amended. Support for the claim amendments may be found in the specification and claims as originally filed. For example, support for the claim amendments may be found in the specification at least at page 3, lines 4-10 and page 5, line 32 – page 6, line 4. No new matter has been added.

Upon entry of this amendment, claims 1-34 will be pending in the present application, with claims 1, 16, 17, 22, 25 and 27 being independent.

Applicants thank Examiners Alam and Johnson for the courtesies extended to applicants' representative, Mr. Sung Kim, during an interview conducted on March 1, 2007. The substance of the interview is incorporated in the remarks that follow.

1) CLAIM OBJECTIONS

Claims 1, 15, 17 and 27 are objected to because of informalities that imply intended use. Claims 1, 15, 17 and 27 have been amended as indicated in the Office Action on page 2.

Claim 13 is objected to because of grammatical errors in the claim. Claim 13 has been amended as indicated in the Office Action on page 2.

Claim 29 is objected to because of informalities. Claim 29 has been amended to clarify that the collection of records is stored in a database relation.

For at least the reasons above, reconsideration and withdrawal of the objections to claims 1, 13, 15, 17, 27 and 29 are respectfully requested.

2) REJECTIONS UNDER 35 U.S.C. 112

Claim 1 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter of the invention. Applicants respectfully traverse this rejection for at least the following reasons.

The Office Action on page 3 asserts that evaluating an input string is never realized in the body of claim 1 and, therefore, there is no nexus between the preamble and the body of the claim. Applicants respectfully disagree.

Claim 1 recites the following element:

> determining a most probable segmentation of the input string <u>by comparing tokens that make up the input string with a state transition model</u> derived from the collection of data records (emphasis added)

Therefore, claim 1 includes the elements of comparing tokens that make up the input string with a state transition model. By comparing the tokens in the input string, the input string is being evaluated. Consequently, evaluating an input string, or a process to evaluate an input string, is realized in the body of claim 1.

For at least the reasons above, reconsideration and withdrawal of the rejection of claim 1 under 35 U.S.C. §112, second paragraph, are respectfully requested.

3) REJECTIONS UNDER 35 U.S.C. 101

Claims 1, 16, 17, 22 and 27 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter. Applicants respectfully traverse this rejection for at least the following reasons.

The Office Action on page 3 asserts that in claims 1, 16 and 27, the act of determining does not produce any functional change, nor does it produce any useful, concrete, and tangible result. As a result, the Office Action asserts that claims 1, 16 and 27 are non-statutory. Applicants disagree with these assertions. However, for purposes of economy of prosecution, claims 1, 16 and 27 have been amended to remove any uncertainty that these claims are directed to statutory subject matter. For example, claim 1 has been amended to include "storing the one or more component parts in a database". Claims 16 and 27 have been amended to include similar elements. Therefore, claims 1, 16 and 27 are directed to statutory subject matter.

The Office Action asserts on pages 3-4 that claims 1, 16, 17, 22 and 27 are directed to program products, which the Examiner deems to be functional descriptive material. The Office Action asserts that the content of these claims is not structurally and functionally

interrelated to a computer-readable medium, thereby rendering the claims incapable of producing a useful, concrete and tangible result. The Office Action also asserts that these claims should be amended to recite hardware in the body of the claims. Applicants disagree with these assertions.

For example, claim 17 has been amended to recite a <u>computer</u> system. Additionally, claim 22 has been amended to recite a string segmentation schema <u>implemented on a computer system</u>. Furthermore, as discussed above, claims 1, 16 and 27 have been amended, as discussed above, to include, in some form, the element of storing one or more component parts in a database. Therefore, claims 1, 16, 17, 22 and 27 are directed to statutory subject matter.

For at least the reasons above, reconsideration and withdrawal of the rejection of claims 1, 16, 17, 22 and 27 under 35 U.S.C. §101 are respectfully requested.

4) REJECTIONS UNDER 35 U.S.C. 102

Claims 1-13, 15, 17-19, 21, 22 and 24-34 are rejected under 35 U.S.C. 102(b) as anticipated by Borkar et al. ("Automatic segmentation of text strings into structured records"). Applicants respectfully traverse this rejection for at least the following reasons.

As discussed during the interview, the approach disclosed in Borkar et al. constitutes a <u>supervised</u> model-based approach for text segmentation in which scalability is achieved by automatically learning segmentation models from <u>manually tagged or segmented training data</u>. An inherent limitation in supervised model-based approaches is that it is often difficult to obtain sufficient training data, especially data that is comprehensive enough to illustrate all features of test data (e.g., see Borkar et al., page 10, section 3.5, 1st paragraph). Furthermore, hand-tagged or segmented training data typically used in supervised model-based approaches, and specifically used by the tool disclosed in Borkar et al., suffers from limitations on the size of training data sets due to the inherently slow and time-consuming human labeling phase in its preparation (e.g., see Borkar et al., page 11, section 5, Acknowledgements, where the authors acknowledge contributors who "painstakingly hand-tagged the test data").

In contrast to Borkar et al., embodiments of the present application are directed to unsupervised text segmentation utilizing a reference table or relation that does not require explicitly labeled (i.e., segmented) training data while building accurate and robust data models for segmenting input strings into structured records (see specification, page 3, lines 2-6).

Specifically, Borkar et al. discloses a tool (DATAMOLD) that learns to automatically extract structure using a Hidden Markov Model when seeded with a small number of training examples (see abstract). The input to DATAMOLD is a fixed set of E elements of the form "House #", "Street" and "City" and a collection of T example addresses that have been segmented into one or more of the elements (see page 3, section 2, 1st paragraph; Figure 1). The collection of T example addresses constitutes training data that is first manually segmented into its constituent elements (see page 8, section 3.1, 5th paragraph). Borkar et al. discloses that the size of the training data is an important concern in all extraction tasks that require manual effort in tagging instances. In most such information extraction problems, untagged data is plentiful but tagged data to serve as training records is scarce and requires human effort (see page 10, section 3.5, 1st paragraph). Therefore, as discussed during the interview, Borkar et al. discloses the use of manually segmented training data to train a Hidden Markov Model to segment input data (see page 3, section 1.3.1, 1st paragraph; page 4, section 2.1, 4th paragraph).

However, Borkar et al. fails to disclose or suggest the elements of providing a state transition model based on an existing collection of data records, wherein the existing collection of data records does not comprise manually segmented training data, as included in amended independent claims 1, 17, 25 and 27. Independent claims 16 and 22 have been amended to include similar elements.

Additionally, the Office Action on pages 4-5 asserts that Borkar et al. discloses the elements of a state transition model based on an existing collection of data records that includes probabilities to segment input strings into component parts which adjusts said probabilities to account for erroneous token placement in the input string (citing page 3, section 1.3.1, lines 19-21). However, the section in Borkar et al. cited by the Office Action merely states:

> The training data helps learn this distribution. During testing, the HMM outputs the most probable state transitions that could have generated an output sequence.

There is no disclosure or suggestion in the above cited section, or elsewhere, in Borkar et al. of the elements of a state transition model based on an existing collection of data records that includes probabilities to segment input strings into component parts which adjusts said probabilities to account for erroneous token placement in the input string, as included, in some form, in independent claims 1, 17 and 27.

Therefore, since Borkar et al. fails to disclose or suggest all of the elements of claims 1, 16, 17, 22, 25 and 27, these claims are allowable over Borkar et al.

Claims 2-13 and 15 depend on claim 1. Claims 18-19 and 21 depend on claim 17. Claim 24 depends on claim 22. Claim 26 depends on claim 25. Claims 28-34 depend on claim 27. As discussed above, claims 1, 17, 22, 25 and 27 are allowable. For at least this reason, and the additional features recited therein, claims 2-13, 15, 18-19, 21, 24, 26 and 28-34 are also allowable.

The Office Action on page 4 asserts that claim 29 is rejected under 35 U.S.C. 102(b) as anticipated by Borkar et al. However, the Office Action fails to specifically address how Borkar et al. anticipates the elements of claim 29. Applicants respectfully request examination of claim 29 on its individual merits.

For at least the reasons above, reconsideration and withdrawal of the rejection of claims 1-13, 15, 17-19, 21, 22 and 24-34 under 35 U.S.C. §102(b) are respectfully requested.

5) REJECTIONS UNDER 35 U.S.C. 103

Claims 14, 20 and 23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Borkar et al. in view of Reed (U.S. Pat. No. 5,095,432). Applicants respectfully traverse this rejection for at least the following reasons.

As discussed above, Borkar et al. fails to disclose or suggest all of the elements of independent claims 1, 17 and 22. Reed fails to cure this defect.

Reed discloses a context-free parsing algorithm employing register vector grammars providing fast parsing of natural languages (see abstract). However, Reed fails to

disclose or suggest the elements of providing a state transition model based on an existing collection of data records, wherein the existing collection of data records <u>does not comprise manually segmented training data</u>, as included, in some form, in independent claims 1, 17 and 22. Furthermore, Reed fails to disclose or suggest the elements of a state transition model based on an existing collection of data records that includes probabilities to segment input strings into component parts which <u>adjusts said probabilities to account for erroneous token placement in the input string</u>, as included, in some form, in independent claims 1 and 17. Therefore, since Borkar et al. and Reed, alone or in combination, fail to disclose or suggest all of the elements of claims 1, 17 and 22, these claims are allowable.

Claim 14 depends on claim 1. Claim 20 depends on claim 17. Claim 23 depends on claim 22. As discussed above, claims 1, 17 and 22 are allowable. For at least this reason, and the additional features recited therein, claims 14, 20 and 23 are also allowable.

Claim 16 is rejected under 35 U.S.C. 103(a) as being unpatentable over Borkar et al. in view of Fairweather (U.S. Pat. App. Pub. No. 2006/0235811). Applicants respectfully traverse this rejection for at least the following reasons.

As discussed above, Borkar et al. discloses the use of <u>manually segmented training data</u> to train a Hidden Markov Model to segment input data (see page 3, section 1.3.1, 1st paragraph; page 4, section 2.1, 4th paragraph). However, Borkar et al. fails to disclose or suggest the elements of providing a reference table of string records that are segmented into multiple substrings corresponding to database attributes, wherein the reference table of string records <u>does not comprise manually segmented training data</u>, as included in claim 16. Fairweather fails to cure this defect.

Fairweather discloses extracting data that produces a strongly-typed ontology defined collection referencing all extracted records (see abstract). However, Fairweather fails to disclose or suggest the elements of providing a reference table of string records that are segmented into multiple substrings corresponding to database attributes, wherein the reference table of string records <u>does not comprise manually segmented training data</u>, as included in claim 16. Therefore, since Borkar et al. and Fairweather, alone or in

combination, fail to disclose or suggest all of the elements of claim 16, this claim is allowable.

For at least the reasons above, reconsideration and withdrawal of the rejection of claims 14, 16, 20 and 23 under 35 U.S.C. §103(a) are respectfully requested.

6) **CONCLUSION**

Accordingly, in view of the above amendments and remarks it is submitted that the claims are patentably distinct over the prior art and that all the rejections to the claims have been overcome. Reconsideration and reexamination of the present application is requested. Based on the foregoing, applicants respectfully request that the pending claims be allowed, and that a timely Notice of Allowance be issued in this case. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call the applicants' attorney at the telephone number listed below.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, applicants hereby request any necessary extension of time. If there is a fee occasioned by this response, including an extension fee that is not covered by an enclosed check please charge any deficiency to Deposit Account No. 50-0463.

Respectfully submitted,

Microsoft Corporation

Date: __March 5, 2007_____   By: ___/Sung T. Kim/_____

Sung T. Kim, Reg. No.: 45,398
Attorney for Applicants
Direct telephone: (703) 647-6574
Microsoft Corporation
One Microsoft Way
Redmond WA 98052-6399

CERTIFICATE OF MAILING OR TRANSMISSION [37 CFR 1.8(a)]

I hereby certify that this correspondence is being electronically deposited with the USPTO via EFS-Web on the date shown below:

_____March 5, 2007_____   _____/Kate Marochkina/_____
Date   Signature

_____Kate Marochkina_____
Type or Print Name